

Metrics in the Space of High Order Networks

WeiYu Huang and Alejandro Ribeiro

Abstract—This paper presents methods to compare high order networks, defined as weighted complete hypergraphs collecting relationship functions between elements of tuples. They can be considered as generalizations of conventional networks where only relationship functions between pairs are defined. Important properties between relationships of tuples of different lengths are established, particularly when relationships encode dissimilarities or proximities between nodes. Two families of distances are then introduced in the space of high order networks. The distances measure differences between networks. We prove that they are valid metrics in the spaces of high order dissimilarity and proximity networks modulo permutation isomorphisms. Practical implications are explored by comparing the coauthorship networks of two popular signal processing researchers. The metrics succeed in identifying their respective collaboration patterns.

I. INTRODUCTION

We consider high order networks that describe relationships between elements of tuples and address the problem of constructing valid metric distances between them. Most often, networks are defined as structures that describe interactions between pairs of nodes [2], [3]. This is an indisputable appropriate model for networks that describe binary relationships, such as communication or influence, but not so appropriate for problems in which binary, ternary, or n -ary relationships in general, have different implications. This is, e.g., true of coauthorship networks where we count the number of joint publications by groups of scholars. Papers written by pairs of authors capture information that can be used to identify important authors and study mores of research communities. However, there is extra information to be gleaned from collaborations between triplets of authors, or even single author publications. The importance of capturing tuple proximities between groups of nodes other than pairs has been recognized and exploited in multiple domains including coverage analysis in sensor networks [4]–[6], cognitive learning and memory [7], broadcasting in wireless networks [8], image ranking [9], three-dimensional object retrieval and recognition [10], and group relationship structure in social networks [11].

The problem of defining distances between networks, or, more loosely, the problem of determining if two networks are similar or not, is important even in the case of pairwise networks. The problem is not complicated if nodes have equal labels in both networks [12]–[15]. The problem, however, becomes very challenging if a common labeling doesn't exist in both networks, as we need to consider all possible mappings between nodes of each network. This complexity has motivated the use of network features as

alternatives to the use of distances. Examples of features that have proved useful in particular settings are clustering coefficients [16], neighborhood topology [17], betweenness [18], motifs [19], wavelets [20], as well as graphlet degree distributions or signatures [21]–[23]. Although feature analysis is often effective, it is application-dependent, utilizes only a small portion of the information conveyed by the networks, and networks not isomorphic may still have zero dissimilarity as measured by features. These drawbacks can be overcome with the definition of valid metric distances that are universal, depend on all edge weights, and are null if and only if the networks are isomorphic [24]. We point out that one can think of defining distances between networks as a generalization of the graph isomorphism problem [25] where the question asked is whether two networks are the same or not. When defining network distances we also want a measure of how far the networks are and we want these measures to be symmetric and satisfy the triangle inequality [24].

The main problem addressed in this paper is the construction of metric distances between high order networks. Formal definitions of high order networks are presented (Section III) as a generalization of pairwise networks (Section II). Dissimilarity networks (Section IV) and proximity networks (Section V) are specific high order networks where relationship functions are intended to encode dissimilarities or proximities between members of tuples. Dissimilarity networks are characterized by the order increasing property which states that tuples become more dissimilar when members are added to a group. Proximity networks abide to the order decreasing property which states that tuples becomes less similar when adding nodes to the group. Two families of proper metric distances are then defined in the respective space of dissimilarity (Section IV-A) and proximity (Section V-A) networks modulo permutation isomorphisms. These distances are built as generalizations of the pairwise distances in [24], which are themselves generalizations of the Gromov-Hausdorff distance between metric spaces [26], [27]. The paper also establishes a duality between dissimilarity and proximity networks and the different metrics (Section V-B). We use the proximity network distances defined in the paper to compare the coauthorship networks of two popular signal processing researchers and show that they succeed in discriminating their collaboration patterns (Section VI). As in the case of pairwise networks these distances can be computed only when the number of nodes is small. Ongoing work is focused on the problem of finding bounds on these network distances that are computable in networks with large numbers of nodes.

II. PAIRWISE NETWORKS

Conventionally, a network is defined as a pair $N_X = (X, r_X^1)$, where X is a finite set of nodes and $r_X^1 : X^2 =$

Work supported by NSF CCF-1217963. The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, 200 South 33rd Street, Philadelphia, PA 19104. Email: whuang, aribeiro@seas.upenn.edu. Part of the results in this paper appeared in [1].

$X \times X \rightarrow \mathbb{R}_+$ is a function that may encode similarity or dissimilarity between elements. For points $x, x' \in X$, values of this function are denoted as $r_X^1(x, x')$. We assume that $r_X^1(x, x') = 0$ if and only if $x = x'$ and we further restrict attention to symmetric networks where $r_X^1(x, x') = r_X^1(x', x)$ for all pairs of nodes $x, x' \in X$. The set of all such networks is denoted as \mathcal{N} .

When defining a distance between networks we need to take into consideration that permutations of nodes amount to relabelling nodes and should be considered as same entities. We therefore say that two networks $N_X = (X, r_X^1)$ and $N_Y = (Y, r_Y^1)$ are isomorphic whenever there exists a bijection $\phi : X \rightarrow Y$ such that for all points $x, x' \in X$,

$$r_X^1(x, x') = r_Y^1(\phi(x), \phi(x')). \quad (1)$$

Such a map is called an isometry. Since the map ϕ is bijective, (1) can only be satisfied when X is a permutation of Y . When networks are isomorphic we write $N_X \cong N_Y$. The space of networks where isomorphic networks $N_X \cong N_Y$ are represented by the same element is termed the set of networks modulo isomorphism and denoted by $\mathcal{N} \text{ mod } \cong$. The space $\mathcal{N} \text{ mod } \cong$ can be endowed with a valid metric [24]. The definition of this distance requires introducing the prerequisite notion of correspondence [28, Def. 7.3.17].

Definition 1 A correspondence between two sets X and Y is a subset $C \subseteq X \times Y$ such that $\forall x \in X$, there exists $y \in Y$ such that $(x, y) \in C$ and $\forall y \in Y$ there exists $x \in X$ such that $(x, y) \in C$. The set of all correspondences between X and Y is denoted as $\mathcal{C}(X, Y)$.

A correspondence in the sense of Definition 1 is a map between node sets X and Y so that every element of each set has at least one correspondent in the other set. Correspondences include permutations as particular cases but also allow for the mapping of a single point in X to multiple correspondents in Y or, vice versa. Most importantly, this allows definition of correspondences between networks with different numbers of elements. We can now define the distance between two networks by selecting the correspondence that makes them most similar as we formally define next.

Definition 2 Given two networks $N_X = (X, r_X^1)$ and $N_Y = (Y, r_Y^1)$ and a correspondence C between the node spaces X and Y define the network difference with respect to C as

$$\Gamma_{X,Y}^1(C) := \max_{(x_1, y_1), (x_2, y_2) \in C} |r_X^1(x_1, x_2) - r_Y^1(y_1, y_2)|. \quad (2)$$

The network distance between networks N_X and N_Y is then defined as

$$d_{\mathcal{N}}^1(N_X, N_Y) := \min_{C \in \mathcal{C}(X, Y)} \left\{ \Gamma_{X,Y}^1(C) \right\}. \quad (3)$$

For a given correspondence $C \in \mathcal{C}(X, Y)$ the network difference $\Gamma_{X,Y}^1(C)$ selects the maximum distance difference $|r_X^1(x_1, x_2) - r_Y^1(y_1, y_2)|$ among all pairs of correspondents – we compare $r_X^1(x_1, x_2)$ with $r_Y^1(y_1, y_2)$ when the points x_1 and y_1 , as well as the points x_2 and y_2 , are correspondents. The distance in (3) is defined by selecting the correspondence

that minimizes these maximal differences. The distance in Definition 2 is a proper metric in the space of networks modulo isomorphism. It is nonnegative, symmetric, satisfies the triangle inequality, and is null if and only if the networks are isomorphic [24]. For future reference, the notions of metric and pseudometric are formally stated next.

Definition 3 Given a space \mathcal{S} and an isomorphism \cong , a function $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ is a metric in $\mathcal{S} \text{ mod } \cong$ if for any $a, b, c \in \mathcal{S}$ the function d satisfies:

- (i) **Nonnegativity.** $d(a, b) \geq 0$.
- (ii) **Symmetry.** $d(a, b) = d(b, a)$.
- (iii) **Identity.** $d(a, b) = 0$ if and only if $a \cong b$.
- (iv) **Triangle inequality.** $d(a, b) \leq d(a, c) + d(c, b)$.

The function is a pseudometric in $\mathcal{S} \text{ mod } \cong$ if for any $a, b, c \in \mathcal{S}$ the function d satisfies (i), (ii), (iv), and

- (iii') **Relaxed identity.** $d(a, b) = 0$ if $a \cong b$.

A metric d in $\mathcal{S} \text{ mod } \cong$ gives a proper notion of distance. Since zero distances imply elements being isomorphic, the distance between elements reflects how far they are from being isomorphic. Pseudometrics are relaxed since elements not isomorphic may still have zero distance measured by the pseudometrics. The distance in Definition 2 is a metric in space $\mathcal{N} \text{ mod } \cong$. Observe that since correspondences may be between networks with different number of elements, Definition 2 defines a distance $d_{\mathcal{N}}^1(N_X, N_Y)$ when the node cardinalities $|X|$ and $|Y|$ are different. In the particular case when the functions r_X^1 satisfy the triangle inequality, the set of networks \mathcal{N} reduces to the set of metric spaces \mathcal{M} . In this case the metric in Definition 2 reduces to the Gromov-Hausdorff (GH) distance between metric spaces. The distances $d_{\mathcal{N}}^1(N_X, N_Y)$ in (3) are valid metrics even if the triangle inequalities are violated by r_X^1 or r_Y^1 [24].

In this paper we consider high order networks where the specification of functions $r_X^k : X^{k+1} \rightarrow \mathbb{R}_+$ are meant to encode similarities or dissimilarities between node $(k+1)$ -tuples. The goal of this paper is to devise generalizations of Definition 2 to high order networks and to prove that they define valid metrics in the space of high order networks modulo isomorphism; see Definitions 11, 12, 14, and 15.

III. HIGH ORDER NETWORKS

A network of order K over the node space X is defined as a collection of $K+1$ relationship functions $\{r_X^k : X^{k+1} \rightarrow \mathbb{R}_+\}_{k=0}^K$ from the space X^{k+1} of $(k+1)$ -tuples to the nonnegative reals,

$$N_X^K = (X, r_X^0, r_X^1, \dots, r_X^K). \quad (4)$$

A network of order K can be considered as a weighted complete hypergraph [29], [30] whose weights for all hyperedges of elements of all $(k+1)$ tuples with $0 \leq k \leq K$ are defined.

When some nodes are repeated in the point collection $x_{0:k} := (x_0, x_1, \dots, x_k) \in X^{k+1}$, the relationship function $r_X^k(x_{0:k})$ entails the same information as the relationship

function between the largest non-repeating subtuple of $x_{0:k}$. In future definitions, it would be important to take the number of distinct elements of a tuple into consideration. We formalize this property by introducing the notion of the rank of tuples as we formally specify next.

Definition 4 The rank $s(x_{0:k})$ of a given tuple $x_{0:k}$ is the number of unique elements in the tuple.

It follows from Definition 4 that the rank $s(x, x) = 1$ and that the rank $s(x', x, x') = 2$. Moreover, the relationship function between a tuple $x_{0:k}$ is identical to the relationship functions of subtuples of $x_{0:k}$ that have same rank as $s(x_{0:k})$ since they imply same information. This remark along with a symmetry property makes up the formal definition of high order networks that we introduce next.

Definition 5 $N_X^K = (X, r_X^0, r_X^1, \dots, r_X^K)$ is a K -order network if the following two properties holds:

Symmetry. For any $0 \leq k \leq K$ and any point collections $x_{0:k}$, we have that

$$r_X^k(x_{[0:k]}) = r_X^k(x_{0:k}), \quad (5)$$

where $x_{[0:k]} = ([x_0], [x_1], \dots, [x_k])$ is a reordering of $x_{0:k} := (x_0, x_1, \dots, x_k)$.

Identity. For any $0 \leq k \leq K$ and tuple $x_{0:k}$, any of its subtuple $x_{l_0:l_k}$ with $s(x_{0:k}) = s(x_{l_0:l_k})$ satisfies

$$r_X^k(x_{0:k}) = r_X^k(x_{l_0:l_k}). \quad (6)$$

The set of all high order networks of order K is denoted as \mathcal{N}^K .

For point collections $x_{0:k}$, values of their k -order relationship functions are denoted as $r_X^k(x_{0:k})$ and are intended to represent a measure of similarity or dissimilarity for members of the group. In particular, the zeroth order function r_X^0 encodes relative weights of different nodes and the first order function r_X^1 represents the pairwise information discussed in Section II. Observe however that pairwise networks are not particular cases of networks of order 1 because a network of order K not only requires the definition of relationships between $(K+1)$ -tuples but also of relationships between $(k+1)$ -tuples for all integers $0 \leq k \leq K$. A network of order 0 is one in which only node weights are given, a network of order 1 is one in which weights and pairwise relationships are defined, a network of order 2 adds relationships between triplets and so on. Examples for the identity property includes $r_X^2(x, x) = r_X^1(x)$ and $r_X^3(x', x, x') = r_X^2(x, x')$. We assume that relationship values are normalized so that $0 \leq r_X^k(x_{0:k}) \leq 1$ for all k and $x_{0:k}$. As in the case of pairwise networks we consider K -order networks N_X^K and N_Y^K to be equivalent for their k -order relationship functions if r_X^k is a permutation of r_Y^k as we formally define next.

Definition 6 We say that two networks N_X^K and N_Y^K are k -isomorphic if there exists a bijection $\phi : X \rightarrow Y$ such that for all $x_{0:k} \in X^{k+1}$ we have

$$r_Y^k(\phi(x_{0:k})) = r_X^k(x_{0:k}), \quad (7)$$

where we use the shorthand notation $r_Y^k(\phi(x_{0:k})) := r_Y^k(\phi(x_0), \phi(x_1), \dots, \phi(x_k))$. The map ϕ is called a k -isometry.

When networks N_X^K and N_Y^K are k -isomorphic we write $N_X^K \cong_k N_Y^K$. The space of K -order networks modulo k -isomorphism is denoted by $\mathcal{N}^K \text{ mod } \cong_k$. For each nonnegative integer $0 \leq k \leq K$, the space $\mathcal{N}^K \text{ mod } \cong_k$ of networks of order K modulo k -isomorphism can be endowed with a pseudometric. The definition of this family of pseudometrics is a generalization of Definition 2 as we formally state next.

Definition 7 Given networks N_X^K and N_Y^K , a correspondence C between the node spaces X and Y , and an integer $0 \leq k \leq K$ define the k -order network difference with respect to C as

$$\Gamma_{X,Y}^k(C) := \max_{(x_{0:k}, y_{0:k}) \in C} |r_X^k(x_{0:k}) - r_Y^k(y_{0:k})|, \quad (8)$$

where the notation $(x_{0:k}, y_{0:k})$ stands for $(x_0, y_0), (x_1, y_1), \dots, (x_k, y_k)$. The k -order network distance between networks N_X^K and N_Y^K is then defined as

$$d_{\mathcal{N}}^k(N_X^K, N_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \{\Gamma_{X,Y}^k(C)\}. \quad (9)$$

We further define the K -order distance vector as the $K+1$ dimensional vector $\mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K) = [d_{\mathcal{N}}^0(N_X^K, N_Y^K), \dots, d_{\mathcal{N}}^K(N_X^K, N_Y^K)]^T$ that groups the k -order distances in (9).

Both, Definition 2 and Definition 7 consider correspondences C that map the node space X onto the node space Y , compare dissimilarities, and set the network distance to the comparison that yields the smallest value in terms of maximum differences. The distinction between them is that in (2) we compare the values in $r_X^1(x_1, x_2)$ and $r_Y^1(y_1, y_2)$, whereas in (8) we compare the values in each of the k -order relationships $r_X^k(x_{0:k})$ and $r_Y^k(y_{0:k})$ to compute the k -order distances $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ that we group in the vector $\mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K)$. Except for this distinction, Definition 2 and Definition 7 are analogous since $\Gamma_{X,Y}^k(C)$ selects the maximum k -order relationship difference $|r_X^k(x_{0:k}) - r_Y^k(y_{0:k})|$ among all tuples of correspondents – we compare $r_X^k(x_{0:k})$ with $r_Y^k(y_{0:k})$ when all the points $x_l \in x_{0:k}$ and $y_l \in y_{0:k}$ are correspondents. The distance $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ is defined by selecting the correspondence that minimizes these maximal differences.

Notice that, in general, the correspondence C minimizing $\Gamma_{X,Y}^k(C)$ is not necessarily identical to the correspondence C' minimizing $\Gamma_{X,Y}^l(C')$ for $k \neq l$. The distance vector $\mathbf{d}_{\mathcal{N}}^K$ is a vector with each element measuring the dissimilarity between relationship functions of a specific order, possibly using different minimizing correspondences. We emphasize that, as in the case of Definition 2, $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ and $\mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K)$ are defined even if the numbers of nodes in X and Y are different. We show in the following proposition that the function $d_{\mathcal{N}}^k : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ is, indeed, a pseudometric in the space of K -order networks modulo k -isomorphism for any integer $0 \leq k \leq K$.

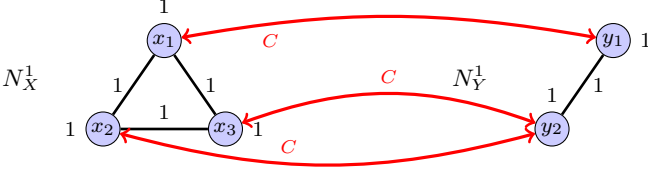


Fig. 1. An example of two networks being not 1-isomorphic but having zero 1-order network distance between them. For the given correspondence C , $r_X^2(x_1, x_2) = r_Y^2(y_1, y_2)$, $r_X^2(x_1, x_3) = r_Y^2(y_1, y_2)$. $r_X^2(x_2, x_3) = r_Y^2(y_2, y_2) = r_Y^1(y_2)$ where the second equality follows from the identity property. Moreover, $r_X^2(x_1, x_1) = r_Y^2(y_1, y_1)$, $r_X^2(x_2, x_2) = r_Y^2(y_2, y_2)$, $r_X^2(x_3, x_3) = r_Y^2(y_2, y_2)$. $\Gamma_{X,Y}^1(C) = 0$ witnesses the zero 1-order network distance between N_X^1 and N_Y^1 . However these networks cannot be 1-isomorphic since they possess different number of nodes.

Proposition 1 Given any nonnegative integer K , for any integers $0 \leq k \leq K$, the function $d_N^k : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ defined in (9) is a pseudometric in the space $\mathcal{N}^K \text{ mod } \cong_k$.

Proof: See Appendix A. ■

d_N^k being a pseudometric implies that two high order networks not k -isomorphic may still have zero k -order network distance between them. A specific example can be found in Figure 1 where two 1-order networks not 1-isomorphic have zero dissimilarity measured by the 1-order network distance. For each integer $0 \leq k \leq K$, the pseudometric $d_N^k(N_X^K, N_Y^K)$ defined in Definition 7 in the space $\mathcal{N}^K \text{ mod } \cong_k$ measures dissimilarity between k -order functions r_X^k and r_Y^k . We can also ask the question of how different two networks are by considering *all* their order functions. To that end we consider K -order networks to be equivalent if r_X^k is a permutation of r_Y^k for all integers $0 \leq k \leq K$ as we formally state next.

Definition 8 We say that two networks of order K , N_X^K and N_Y^K , are isomorphic if there exists a bijection $\phi : X \rightarrow Y$ such that (7) holds for all $0 \leq k \leq K$ and $x_{0:k} \in X^{k+1}$. The map ϕ is called an isometry.

When networks N_X^K and N_Y^K are isomorphic we write $N_X^K \cong N_Y^K$. The difference between k -isomorphism and isomorphism is that the bijection in the latter case preserves relationship functions over all orders whereas only k -order relationship functions are preserved in the former case. That $N_X^K \cong N_Y^K$ implies that $N_X^K \cong_k N_Y^K$ for all integers $0 \leq k \leq K$, but the opposite is not necessarily true.

The space of K -order networks modulo isomorphism is denoted as $\mathcal{N}^K \text{ mod } \cong$. A family of pseudometrics measuring the difference between networks over all order functions as a whole can be endowed in the space $\mathcal{N}^K \text{ mod } \cong$. The definition of this family of distances can be considered as an extension of Definition 2 and an aggregation of Definition 7 as we formally state next.

Definition 9 Given networks N_X^K and N_Y^K , a correspondence C between the node spaces X and Y , and some p -norm $\|\cdot\|_p$,

define the network difference with respect to C as

$$\|\Gamma_{X,Y}^K(C)\|_p := \left\| \left(\Gamma_{X,Y}^0(C), \Gamma_{X,Y}^1(C), \dots, \Gamma_{X,Y}^K(C) \right)^T \right\|_p, \quad (10)$$

where for each integer $0 \leq k \leq K$, $\Gamma_{X,Y}^k(C)$ is the k -order network difference with respect to C defined in (8). The p -norm network distance between N_X^K and N_Y^K is then defined as

$$d_{N,p}(N_X^K, N_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \left\{ \|\Gamma_{X,Y}^K(C)\|_p \right\}. \quad (11)$$

The difference between Definition 2, Definition 7 and Definition 9 is that in the case of the network distance $d_{N,p}(N_X^K, N_Y^K)$, we compare not only relationship functions $r_X^k(x_{0:k})$ and $r_Y^k(y_{0:k})$ but also all the relationship functions of order not larger than K . The norm over the vector $\Gamma_{X,Y}^K(C)$ formed by k -order network differences with respect to C for all integers $0 \leq k \leq K$ is assigned as the difference between N_X^K and N_Y^K measured by the correspondence C . The distance $d_{N,p}(N_X^K, N_Y^K)$ is then defined as the minimum of these differences achieved by some correspondence. As in the cases of Definition 2 and Definition 7, $d_{N,p}(N_X^K, N_Y^K)$ is defined even if the numbers of nodes in X and Y are different. The function $d_{N,p} : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ is a pseudometric in the space of K -order networks modulo isomorphism as we show in the following proposition.

Proposition 2 Given some p -norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{N,p} : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ defined in (11) is a pseudometric in the space $\mathcal{N}^K \text{ mod } \cong$.

Proof: See Appendix B. ■

Observe that in (11) we are only allowed to pick one correspondence minimizing $\|\Gamma_{X,Y}^K(C)\|_p$ whereas in (9) for each k we are able to pick one correspondence minimizing the order specific $\Gamma_{X,Y}^k(C)$. This establishes a relationship between $d_{N,p}$ and $\|\mathbf{d}_N^K\|_p$ that we show next.

Proposition 3 Given some p -norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{N,p}$ defined in (11) is no smaller than $\|\mathbf{d}_N^K\|_p$ where \mathbf{d}_N^K is the vector of distances defined in Definition 7. I.e., for any pair of K -order networks N_X^K, N_Y^K , we have that

$$d_{N,p}(N_X^K, N_Y^K) \geq \|\mathbf{d}_N^K(N_X^K, N_Y^K)\|_p. \quad (12)$$

Proof: Given K -order networks N_X^K, N_Y^K , a correspondence C between the node spaces X and Y , and an integer $0 \leq k \leq K$, it follows from (9) that

$$\Gamma_{X,Y}^k(C) \geq d_N^k(N_X^K, N_Y^K). \quad (13)$$

This implies that the vector $\mathbf{d}_N^K(N_X^K, N_Y^K)$ is element-wise no greater than $\Gamma_{X,Y}^K(C)$ from where it follows that

$$\|\Gamma_{X,Y}^K(C)\|_p \geq \|\mathbf{d}_N^K(N_X^K, N_Y^K)\|_p. \quad (14)$$

Since (14) applies for any correspondence C , the minimum of $\|\Gamma_{X,Y}^K(C)\|_p$ achieved by some correspondence in

the set of correspondence $\mathcal{C}(X, Y)$ is still no smaller than $\|\mathbf{d}_N^K(N_X^K, N_Y^K)\|_p$,

$$\min_{C \in \mathcal{C}(X, Y)} \left\{ \|\Gamma_{X, Y}^K(C)\|_p \right\} \geq \|\mathbf{d}_N^K(N_X^K, N_Y^K)\|_p. \quad (15)$$

The result in (12) follows after noting that the minimum in the left hand side of (15) is the distance $d_{N, p}(N_X^K, N_Y^K)$ in (11). ■

Definitions 7 and 9 are *pseudometrics* in the space of high order networks modulo appropriate isomorphisms. To obtain proper *metrics*, we restrict attention to subclasses of networks having specific structures. To do so, observe that the k -order function r_X^k of a given network N_X^K does not impose constraints on the l -order function r_X^l of the same network except the identity property. In practical situations, however, it is common to observe that adding nodes to a tuple results in either increasing or decreasing relationships between elements of the extended tuple. This motivates the consideration of dissimilarity networks and proximity networks that we undertake in the next two sections.

IV. DISSIMILARITY NETWORKS

In dissimilarity networks the function $r_X^k(x_{0:k})$ encodes a level of dissimilarity between elements of the $x_{0:k}$ tuple. In this scenario it is reasonable to assume that adding elements to a tuple makes the group more dissimilar. This restriction along with a generalization of the requirement that $r_X^1(x, x') = 0$ if and only if $x = x'$ in pairwise network makes up the formal definition that we introduce next.

Definition 10 We say that the K -order network $D_X^K = (X, r_X^0, r_X^1, \dots, r_X^K)$ is a dissimilarity network if for any order $0 \leq k \leq K$ and tuples $x_{0:k} \in X^{k+1}$, its relationship function is the summation of a dissimilarity function and the multiplication of its rank with a small constant ϵ ,

$$r_X^k(x_{0:k}) = d_X^k(x_{0:k}) + \epsilon s(x_{0:k}) \quad (16)$$

The dissimilarity terms satisfy the order increasing property so that for any $1 \leq k \leq K$ and $x_{0:k}$,

$$d_X^k(x_{0:k}) \geq d_X^{k-1}(x_{0:k-1}), \quad (17)$$

and the constant $\epsilon > 0$ is a strictly positive value that satisfies

$$0 < \epsilon \leq 1 - \frac{1}{K} \max_{\tilde{x}_{0:K} \in X^{K+1}} d_X^K(\tilde{x}_{0:K}). \quad (18)$$

The set of all dissimilarity networks of order K is denoted as \mathcal{D}^K .

To see that the order increasing property (17) in Definition 10 is reasonable consider a network describing the temporal dynamics of the formation of a research community – see Figure 2. The dissimilarity term in the k -order relationship function in this network marks the normalized time instant at which members of a given $(k+1)$ -tuple write their first joint paper. In particular, the zeroth order dissimilarities d_X^0 are the normalized time instants when authors publish their first paper. In Figure 2 authors A, B, C , and D publish their first papers at times 0, 1/9, 5/9, and 3/9. The first order

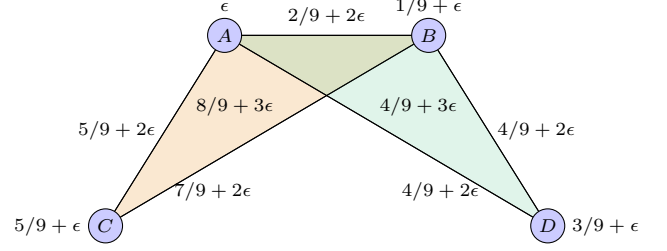


Fig. 2. Temporal dynamics for the formation of a research community. The k -order relationship function in this 2-order dissimilarity network [cf. Definition 10] incorporates the dissimilarity function – the normalized time instant at which members of a given $(k+1)$ -tuple write their first joint paper – and the multiplication of ϵ with the rank of the tuple. E.g., A writes her first paper at time 0, and coauthors with B, D , and C at times 2/9, 4/9, and 5/9. She also writes jointly with B and D at time 4/9.

dissimilarities d_X^1 between pairs denote the normalized times at which nodes become coauthors. Since authors can't become coauthors until after they write their first paper it is certain that $d_X^1(x, x') \geq d_X^0(x)$ and $d_X^1(x, x') \geq d_X^0(x')$ for all x and x' . In Figure 2, A and B become coauthors at time 2/9, which occurs after they publish their respective first papers at times 0 and 1/9. Authors A and D as well as B and D become coauthors at time 4/9, A and C become coauthors at time 5/9. Authors C and D never write a paper together.

Second order dissimilarities d_X^2 for triplets denote the normalized time at which a paper is coauthored by the three members of the triplet. Since a paper can't be coauthored by three people without being at the same time coauthored by each of the three possible pairs of authors we must have that $d_X^2(x, x', x'') \geq d_X^1(x, x')$, $d_X^2(x, x', x'') \geq d_X^1(x, x'')$, and $d_X^2(x, x', x'') \geq d_X^1(x', x'')$ for all x, x' , and x'' . In Figure 2, authors A, B , and D publish a joint paper at time 4/9, which is no smaller than the pairwise coauthorship times between each two of the individual authors. Authors A, B , and C publish a joint paper at time 8/9, which is a time that comes after the individual paired publications that occur at times 2/9, 5/9, and 7/9. Note that due to symmetry property a relationship as in (17) holds if we remove an arbitrary node from the tuple $x_{0:k}$, not necessarily the last.

In pairwise dissimilarity networks we required $d_X^1(x, x') = 0$ if and only if $x = x'$. Relationships between two different nodes are *strictly greater* than relationships between two nodes that are actually identical. The multiplication of ϵ and the rank of the tuples in (16) in Definition 10 can be considered as a generalization. Consider tuples $x_{0:k}$ and $(x_{0:k-1}, x_0)$ where every node in $x_{0:k}$ is unique, the identity property for high order networks forces $r_X^k(x_{0:k-1}, x_0) = r_X^{k-1}(x_{0:k-1})$. We must then have the relationship between $k+1$ different elements $r_X^k(x_{0:k})$ being strictly greater than the relationship between k different elements $r_X^k(x_{0:k-1}, x_0) = r_X^{k-1}(x_{0:k-1})$. This is because $d_X^k(x_{0:k}) \geq d_X^{k-1}(x_{0:k-1})$ follows from (17) and $\epsilon s(x_{0:k}) = (k+1)\epsilon > k\epsilon = \epsilon s(x_{0:k-1})$ follows from the definition of ranks. Therefore, the multiplication of ϵ and the rank of tuples in (16) in Definition 10 forces that adding a new element to a tuple makes the set *strictly more dissimilar* than it was. Or equivalently, removing an element from a tuple makes the set *strictly less dissimilar* than it was. The requirement

for ϵ as in (18) ensures that the highest relationship in the network $\max_{\tilde{x}_{0:K} \in X^{K+1}} d_X^K(\tilde{x}_{0:K}) + \epsilon s(\tilde{x}_{0:K})$ is bounded above by 1. The rank correction term $\epsilon s(x_{0:k})$ is a technical modification to distinguish between full rank (proper) k -tuples and rank deficient (degenerate) tuples. In practice it can be set to a sufficiently small value compared to dissimilarities or completely ignored. Since distances up to order 2 are defined and relationship functions can be decomposed, the network in Figure 2 is a dissimilarity network of order 2.

A. Metrics in the space of dissimilarity networks

When the input networks in Definition 7 are dissimilarity networks we refer to the k -order distance as the k -order dissimilarity network distance. We state this formally in the following definition for future reference.

Definition 11 Given dissimilarity networks $D_X^K, D_Y^K \in \mathcal{D}^K$ we say that the k -order distance $d_N^k(D_X^K, D_Y^K) = d_D^k(D_X^K, D_Y^K)$ of Definition 7 is the k -order dissimilarity network distance between D_X^K and D_Y^K .

Since $\mathcal{D}^K \subseteq \mathcal{N}^K$, the function $d_D^k : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ is a pseudometric in the space of K -order dissimilarity networks modulo k -isomorphism. The restriction, however, makes d_D^k not only a pseudometric but a well-defined metric in the space $\mathcal{D}^K \bmod \cong_k$ of dissimilarity networks of order K modulo k -isomorphism. We show this in the following theorem.

Theorem 1 The k -order dissimilarity network distance function $d_D^k : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ of Definition 11 is a metric in the space $\mathcal{D}^K \bmod \cong_k$ for all $1 \leq k \leq K$.

Proof: See Appendix C. ■

Observe that in Theorem 1 we have that d_D^k is a proper metric for all k other than 0. This caveat for d_D^0 is because we may have two dissimilarity networks D_X^K and D_Y^K with different number of nodes but whose zeroth order relationships are equal for all pairs of nodes, i.e., $r_X^0(x) = r_Y^0(y)$ for all $x \in X$ and $y \in Y$. In this case we would have $d_D^0(D_X^K, D_Y^K) = 0$, however the two dissimilarity networks are not 0-isomorphic.

Restricting Definition 9 to dissimilarity networks also yields a family of dissimilarity network distances as next.

Definition 12 Given dissimilarity networks $D_X^K, D_Y^K \in \mathcal{D}^K$ we say that the p -norm network distance $d_{N,p}(D_X^K, D_Y^K) = d_{D,p}(D_X^K, D_Y^K)$ of Definition 9 is the p -norm dissimilarity network distance between D_X^K and D_Y^K .

By restricting our attention to dissimilarity networks instead of general high order networks, $d_{D,p}$ also becomes a valid metric in the space $\mathcal{D}^K \bmod \cong$ of dissimilarity networks of order $K \geq 1$ modulo isomorphism as we state in the following theorem.

Theorem 2 Given some p -norm $\|\cdot\|_p$, for any nonnegative integer $K \geq 1$ the function $d_{D,p} : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ in Definition 12 is a metric in the space $\mathcal{D}^K \bmod \cong$.

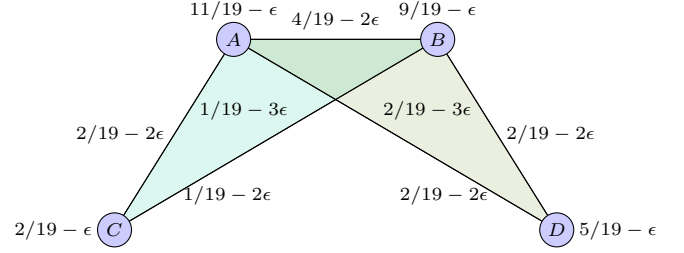


Fig. 3. Collaborations between authors in a research community. The k -order relationship function in this 2-order network [cf. Definition 13] incorporates the proximity function – the number of publications between members of a given $(k+1)$ -tuples normalized by the total number of papers – and the multiplication of $-\epsilon$ with the rank of the tuple.

Proof: See Appendix C. ■

Further note that since Proposition 3 holds for any pair of networks, the same relationship holds true for the dissimilarity network distances in Definitions 11 and 12. Observe, however, that the norm $\|\mathbf{d}_D^k(D_X^K, D_Y^K)\|_p$ is not a valid metric because we can have instances in which two dissimilarity networks are k -isomorphic for all integers $0 \leq k \leq K$ without being isomorphic.

V. PROXIMITY NETWORKS

In proximity networks the relationship functions $r_X^k(x_{0:k})$ denote similarity or proximity between elements of a tuple. Thus, large values of the proximity function $r_X^k(x_{0:k})$ represent strong relationship whereas small values denote weak relationships – the exact opposite is true of dissimilarity networks. In this framework it is reasonable to assume that adding elements to a tuple forces the group to be less similar. This constraint makes up the formal definition we introduce as follows.

Definition 13 We say that the K -order network $P_X^K = (X, r_X^0, r_X^1, \dots, r_X^K)$ is a proximity network if for any order $0 \leq k \leq K$ and tuples $x_{0:k} \in X^{k+1}$, its relationship function is the summation of a proximity term and the multiplication of its rank with $-\epsilon$,

$$r_X^k(x_{0:k}) = d_X^k(x_{0:k}) - \epsilon s(x_{0:k}), \quad (19)$$

The proximity terms satisfy the order increasing property that for any $1 \leq k \leq K$ and $x_{0:k}$,

$$p_X^k(x_{0:k}) \leq p_X^{k-1}(x_{0:k-1}), \quad (20)$$

and the constant $\epsilon > 0$ is a strictly positive value that satisfies

$$0 < \epsilon \leq \frac{1}{K} \min_{\tilde{x} \in X^{K+1}} p_X^K(\tilde{x}_{0:K}). \quad (21)$$

The set of all proximity networks of order K is denoted as \mathcal{P}^K .

To see that the order decreasing property (20) in Definition 13 is reasonable, consider a network illustrating the collaborations between authors in a research community – See Figure 3. The k -order proximity function in this network labels the number of publications between members of a given

$(k+1)$ -tuple. In specific, the zeroth order proximities p_X^0 are the numbers of papers published by authors normalized by the total number of papers. In Figure 3 authors A, B, C, D publish 11, 9, 2, 5 papers respectively and there are 19 papers in total which implies $p_X^0(A) = 11/19$, $p_X^0(B) = 9/19$, $p_X^0(C) = 2/19$, $p_X^0(D) = 5/19$. The first order proximities p_X^1 represent the number of papers co-published by nodes. Since collaboration for a pair of authors is also a paper for each of the individuals it is certain that $p_X^1(x, x') \leq p_X^0(x)$ and $p_X^1(x, x') \leq p_X^0(x')$ for all x and x' . In Figure 3, A and B collaborate on 4 papers, which is less than the 11 and 9 papers written by each of the individuals. Authors A and C as well as A and D coauthor 2 papers in total. Authors C and D never write a paper together.

Second order proximities p_X^2 for triplets indicate the normalized number of papers coauthored by the three members of the triplet. Since a paper with three authors is also a collaboration for the three pairs of authors we must have $p_X^2(x, x', x'') \leq p_X^1(x, x')$, $p_X^2(x, x', x'') \leq p_X^1(x, x'')$, and $p_X^2(x, x', x'') \leq p_X^1(x', x'')$ for all x, x' , and x'' . In Figure 3, authors A, B , and D cowrite 2 papers, which is no more than the number of pairwise collaborations between each pair of the authors. Remark that symmetry property inherited from high order networks [cf. Definition 5] implies (20) if we remove an arbitrary node from the tuple $x_{0:k}$, not necessarily the last.

In dissimilarity networks we required the relationship within tuple $x_{0:k}$ of unique elements to be strictly greater than the relationship between the point collection $(x_{0:k-1}, x_0)$ where some nodes are repeating. The multiplication of $-\epsilon$ and ranks in (19) in Definition 13 can also be considered as a generalization. Following the identity property of high order networks, $r_X^k(x_{0:k-1}, x_0) = r_X^{k-1}(x_{0:k-1})$. We must then have the function between $k+1$ different elements $r_X^k(x_{0:k})$ being strictly smaller than the function between k different elements $r_X^k(x_{0:k-1}, x_0) = r_X^{k-1}(x_{0:k-1})$. This is because in the decomposition $p_X^k(x_{0:k}) \leq p_X^{k-1}(x_{0:k-1})$ follows from (20) and $-\epsilon s(x_{0:k}) = -(k+1)\epsilon < -k\epsilon = -\epsilon s(x_{0:k-1})$ follows from the definition of ranks. Therefore, the multiplication of $-\epsilon$ and rank of tuples in (19) in Definition 10 forces that adding a new element to a tuple makes the set *strictly less similar* than it was. Or equivalently, removing an element from a tuple makes the set strictly more similar than it was. The requirement for ϵ as in (21) ensures that the lowest relationship function in the network $\min_{\tilde{x}_{0:K} \in X^{K+1}} d_X^K(\tilde{x}_{0:K}) - \epsilon s(\tilde{x}_{0:k})$ is nonnegative. Again the rank correction term $\epsilon s(x_{0:k})$ is a technical modification and in practice it can be set to sufficiently small compared to proximities or completely ignored. Since relationships up to order 2 are defined and can be decomposed, the network in Figure 3 is a proximity network of order 2.

A. Metrics in the space of proximity networks

In the same way that restricting attention to dissimilarity networks transforms the pseudometrics in Definitions 7 and 9 into metrics, restricting attention to proximity networks also results in the definitions of proper metrics. We state the restrictions of Definitions 7 and 9 in the following two definitions.

Definition 14 Given proximity networks $P_X^K, P_Y^K \in \mathcal{P}^K$ we say that the k -order distance $d_N^k(P_X^K, P_Y^K) = d_P^k(P_X^K, P_Y^K)$ of Definition 7 is the k -order proximity network distance between P_X^K and P_Y^K .

Definition 15 Given proximity networks $P_X^K, P_Y^K \in \mathcal{P}^K$ we say that the p -norm network distance $d_{N,p}(P_X^K, P_Y^K) = d_{P,p}(P_X^K, P_Y^K)$ of Definition 9 is the p -norm proximity network distance between P_X^K and P_Y^K .

Analogously to the definition of the dissimilarity network distance d_D^k of Definition 11, the function $d_P^k : \mathcal{P}^K \times \mathcal{P}^K \rightarrow \mathbb{R}_+$ is a proper metric in the space $\mathcal{P}^K \bmod \cong_k$ of proximity networks of order K modulo k -isomorphism for all integers $1 \leq k \leq K$. Likewise, restricting the function $d_{N,p}$ of Definition 9 to proximity networks as Definition 15 results in $d_{P,p}$ being a proper metric. We state these facts in the following theorems.

Theorem 3 The k -order proximity network distance function $d_P^k : \mathcal{P}^K \times \mathcal{P}^K \rightarrow \mathbb{R}_+$ of Definition 14 is a metric in the space $\mathcal{P}^K \bmod \cong_k$ for all $k \geq 1$.

Theorem 4 Given some p -norm $\|\cdot\|_p$, for any nonnegative integer $K \geq 1$ the function $d_{P,p} : \mathcal{P}^K \times \mathcal{P}^K \rightarrow \mathbb{R}_+$ in Definition 15 is a metric in the space $\mathcal{P}^K \bmod \cong$.

Proof: See Appendix D. ■

In Theorem 3 we require $k \geq 1$ for the same reason as in Theorem 1. We emphasize that d_P^k is a metric in the space of proximity network modulo k -isomorphisms, whereas $d_{P,p}$ is a metric in the space of networks modulo isomorphism. Also note that we must have $d_{P,p}(P_X^K, P_Y^K) \geq \|\mathbf{d}_P^k(P_X^K, P_Y^K)\|_p$ as per Proposition 3 but $\|\mathbf{d}_P^k(P_X^K, P_Y^K)\|_p$ is not necessarily a metric.

Remark 1 GH distance is the minimum across correspondences of the maximum difference in distances between pairs of nodes for a given correspondence. The metric definitions as in Definitions 11, 12, 14, and 15 inherit this property, which means that network distances can be dominated by a small portion of the networks. Put differently, the proposed distances are more sensitive to a few large differences in a few edges than to a large number of small differences in a large number of edges. Analogous consideration can be found in signal processing theory of the tradeoffs between comparing signals with averages – such as 2-norm comparisons – and comparing signals with max-min differences – the ∞ -norm comparison. When compare networks with different number of nodes, a max-min comparison is reasonable because it focuses attention in the bottleneck tuple that makes it impossible to match smaller network onto the larger.

Remark 2 Once endowed with the proposed valid metrics as in Definitions 11, 12, 14, and 15, the space of dissimilarity networks and the space of proximity networks become metric spaces. This implies that a number of algorithms that are used to analyze metric spaces can now be used to analyze high order networks.

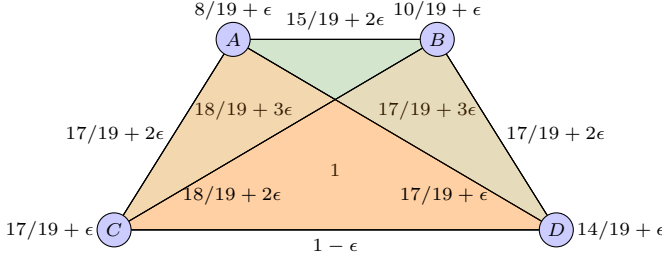


Fig. 4. Relationships between authors expressed in terms of dissimilarities constructed from the proximity network in Figure 3. The k -order relationship function in this 2-order network denotes the level of dissimilarities between members of a given $(k+1)$ -tuples. This is a dissimilarity network that has same order and identical node sets as the proximity network.

B. Duality between dissimilarity and proximity networks

Proximity and dissimilarity networks have been defined separately for simplicity of presentation, but they are actually related entities. For any proximity network P_X^K with relationship functions $\hat{p}_X^k(x_{0:k})$, we can construct a dissimilarity network D_X^K on the same node space by defining relationships as $\hat{d}_X^k(x_{0:k}) = 1 - \hat{p}_X^k(x_{0:k})$ for all orders k and tuples $x_{0:k}$. Likewise given a dissimilarity network D_X^K with relationship functions $\hat{d}_X^k(x_{0:k})$ we can construct a proximity network P_X^K by defining relationships $\hat{p}_X^k(x_{0:k}) = 1 - \hat{d}_X^k(x_{0:k})$. We formalize this equivalence through the introduction of dual networks in the following definition.

Definition 16 Given a node space X , the K -order proximity and dissimilarity networks $P_X^K = (X, \hat{p}_X^0, \hat{p}_X^1, \dots, \hat{p}_X^K)$ and $D_X^K = (X, \hat{d}_X^0, \hat{d}_X^1, \dots, \hat{d}_X^K)$ are said duals if and only if

$$\hat{p}_X^k(x_{0:k}) = 1 - \hat{d}_X^k(x_{0:k}), \quad (22)$$

for all orders $0 \leq k \leq K$ and tuples $x_{0:k}$.

It is ready to see that all proximity networks have a dual dissimilarity network and that, conversely, all dissimilarity networks have a dual proximity network. To do so we just reinterpret (22) as a definition and observe that: (i) The decomposition of relationships in the proximity network implies the valid decomposition of relationships in the dual dissimilarity network, and vice versa. (ii) The order decreasing property of the proximities in the proximity network implies the order increasing property of the dissimilarities in the dual dissimilarity network, and vice versa. An illustration for the construction of a dual dissimilarity network is presented in Figure 4, where we construct the corresponding dual dissimilarity network for the coauthorship network considered in Figure 3.

Given dual networks we can compute the distances in definitions 14 and 15 for proximity networks and the distances in definitions 11 and 12 for the dual dissimilarity networks. These definitions have been constructed so that the resulting distances are the same, as we formally state in the following proposition.

Proposition 4 Consider two proximity networks P_X^K and P_Y^K and their corresponding dual dissimilarity networks D_X^K and

D_Y^K . The k -order proximity distances $d_P^k(P_X^K, P_Y^K)$ [cf. Definition 14] and k -order dissimilarity distances $d_D^k(D_X^K, D_Y^K)$ [cf. Definition 11] coincide for all $0 \leq k \leq K$,

$$d_P^k(P_X^K, P_Y^K) = d_D^k(D_X^K, D_Y^K). \quad (23)$$

Likewise, the p -norm proximity distance $d_{P,p}(P_X^K, P_Y^K)$ [cf. Definition 15] and p -norm dissimilarity distance $d_{D,p}(D_X^K, D_Y^K)$ [cf. Definition 12] coincide,

$$d_{P,p}(P_X^K, P_Y^K) = d_{D,p}(D_X^K, D_Y^K). \quad (24)$$

Proof: See Appendix E. ■

VI. COMPARISON OF COAUTHORSHIP NETWORKS

We apply the metrics defined in Section V-A to compare second order coauthorship networks where relationship functions denote the number of publications of single authors, pairs of authors, and triplets. These coauthorship networks are proximity networks because they satisfy the order decreasing property in Definition 13. Since both, Definition 14 and Definition 15, require searching over all possible correspondences between the node spaces, we can compute exact distances for networks with a small number of nodes only. Thus, we consider publications in the IEEE Transactions on Signal Processing (TSP) in the last decade but restrict attention to the collaboration networks of Prof. Georgios B. Giannakis (GG) of the University of Minnesota and Prof. Martin Vetterli (MV) of the École Polytechnique Fédérale de Lausanne. We choose these authors because their collaboration traits are more developed and stable and we expect their respective collaboration pattern to be steady over the past decade. The goal of the simulation is to illustrate that network metrics are able to distinguish discernible collaboration patterns. For each of the authors, GG and MV, we construct networks for the 2004-2008 and 2009-2013 quinquennia. These networks are referred as GG0408, GG0913, MV0408, and MV0913. For GG we also define networks for each of the biennia 2004-2005, 2006-2007, 2008-2009, 2010-2011, and 2012-2013. We denote these networks as GG0405, GG0607, GG0809, GG1011, and GG1213. Lists of publications are queried from [31].

For each of these authors we consider all of their TSP publications in the period of interest and construct proximity networks where the node space X is formed by the author and the respective set of coauthors. Zeroth order proximities are defined as the total number of publications of each member of the network, first order proximities as the number of papers coauthored by pairs, and second order proximities as the number of papers coauthored by triplets. The constant ϵ as in Definition 13 is for technical purpose. It can be chosen sufficiently small and for this reason we ignore it in this section. To make networks with different numbers of papers comparable we normalize all distances by the total number of papers in the network. With this construction we have that the zeroth order proximity of GG or MV are 1 in all of their respective networks. There are papers with more than three coauthors but we don't record proximities of order higher than 2.

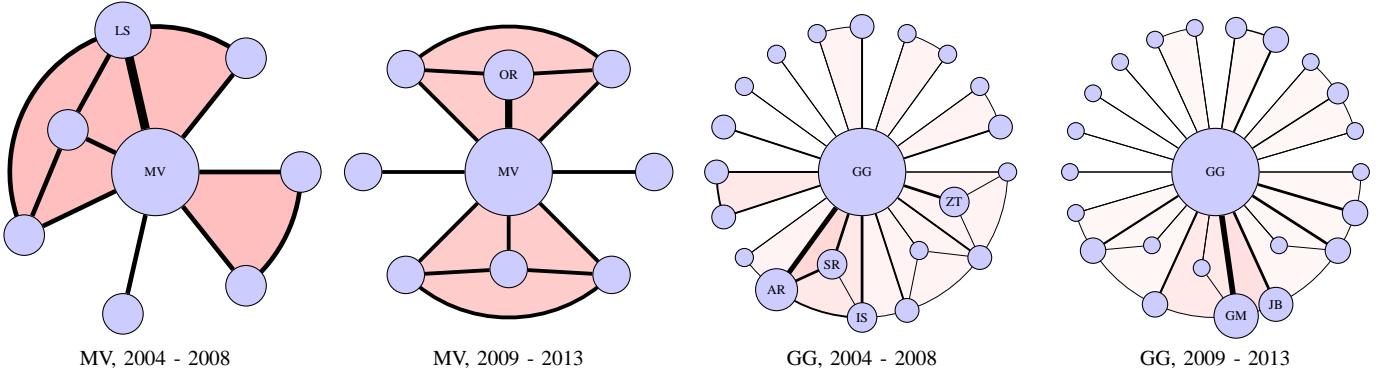


Fig. 5. Quinquennial coauthorship networks representing research communities centered at Prof. Georgios Giannakis (GG) or Prof. Martin Vetterli (MV). The size of the nodes is proportional to the zeroth order proximities, and the width of the links to the first order proximities. Second order proximities are represented by shading the triangle enclosed by the coauthor triplet. Color intensity is proportional to the second order proximities.

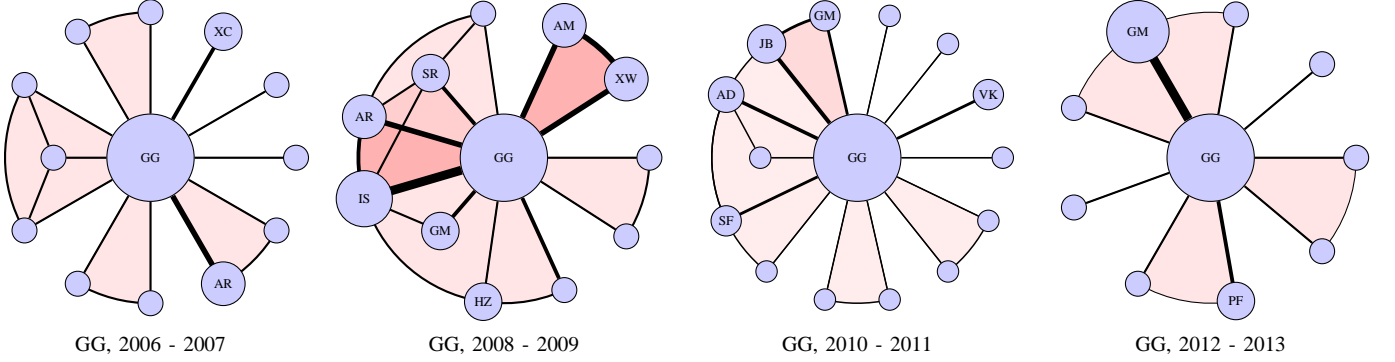


Fig. 6. Biennial coauthorship networks representing research communities centered at Prof. Georgios Giannakis (GG).

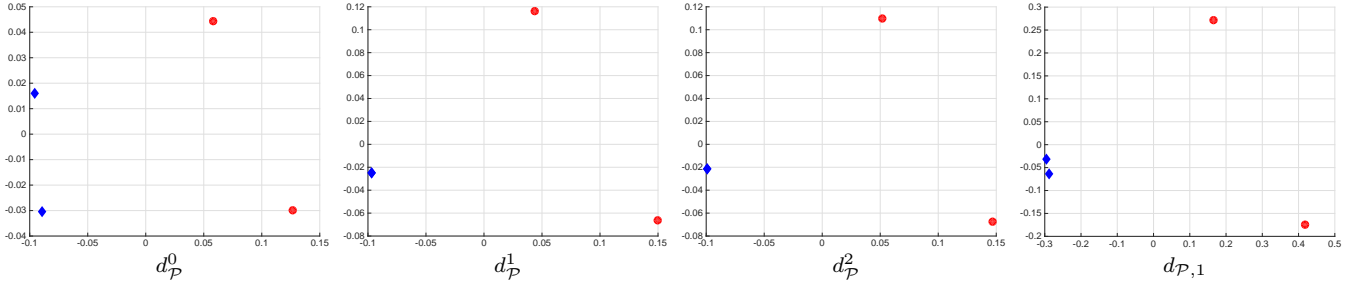


Fig. 7. Two dimensional Euclidean embeddings of the k -order proximity network distances d_P^0, d_P^1, d_P^2 and the proximity network distance with respect to the 1-norm, $d_{P,1}$, between the quinquennial networks. In the embeddings, denote MV0408, MV0913 as circles, GG0408, GG0913 as diamonds. GG0408 and GG0913 are collocated regarding d_P^0, d_P^2 .

The quinquennial networks GG0408, GG0913, MV0408, and MV0913 are shown in Figure 5 and the biennial networks GG0607, GG0809, GG1011, and GG1213 in Figure 6. The size of the nodes is proportional to the zeroth order distances, and the width of the links to the first order distances. Second order proximities are represented by shading the triangle enclosed by the coauthor triplet and the color intensity is proportional to the second order proximities. There are clear differences in the collaboration patterns. We show here that proximity network distances succeed in identifying these patterns and distinguish between the coauthorship networks of GG and MV.

A. Quinquennial networks

Two dimensional Euclidean embeddings (respect to minimizing the sum of squares of the interpoint distances) of the

k -order proximity network distances d_P^k for $k \in \{0, 1, 2\}$ and the proximity network distance with respect to the 1-norm, $d_{P,1}$ are shown in Figure 7. The two GG networks (diamonds) separate clearly from the two MV networks (circles) either by considering the individual k -order distances d_P^k or the aggregate distance $d_{P,1}$. The distances between the two MV networks are high but still smaller than the distances between GG networks and MV networks. An unsupervised classification run across all four distances would assign all four networks correctly.

The k -order network distance d_P^k is defined by searching for the correspondence such that the maximum k -order proximity difference $|r_X^k(x_{0:k}) - r_Y^k(y_{0:k})|$ among all tuples of correspondents is minimized [cf. (8) and (9)]. For the optimal correspondence $C^* = \operatorname{argmin}_{C \in \mathcal{C}(X,Y)} \Gamma_{X,Y}^k(C)$, define the

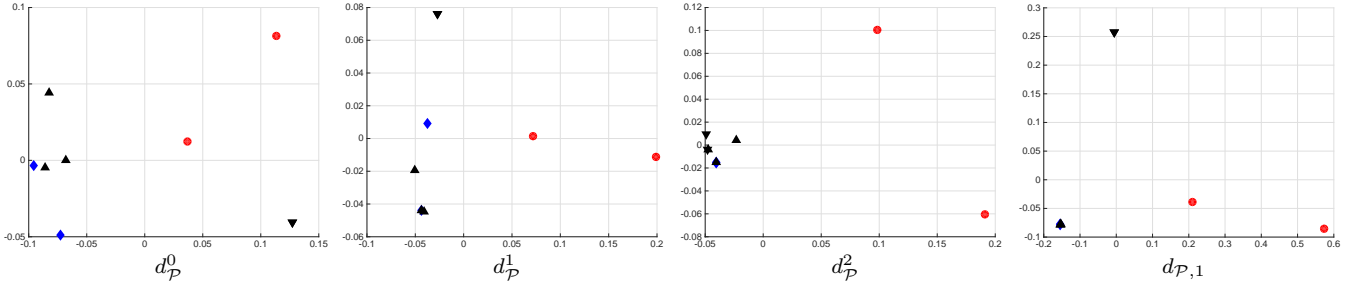


Fig. 8. Two dimensional Euclidean embeddings of the distances $d_P^0, d_P^1, d_P^2, d_{P,1}$ between all quinquennial and biennial networks. In the embeddings, denote MV0408, MV0913 as circles, GG0408, GG0913 as diamonds, GG0405, GG0607, GG1011 as up triangles and GG0809, GG1213 as down triangles. GG0809, GG1213 are colocated regarding $d_P^0, d_P^1, d_{P,1}$. GG0408 and GG0913 have identical coordinates in d_P^1 .

pair of correspondent tuples that achieve the maximum k -order difference as

$$(x_{0:k}^*, y_{0:k}^*) = \underset{(x_{0:k}, y_{0:k}) \in C^*}{\operatorname{argmax}} |r_X^k(x_{0:k}) - r_Y^k(y_{0:k})|. \quad (25)$$

The tuple pair $(x_{0:k}^*, y_{0:k}^*)$ is the bottleneck that prevents making the networks closer to each other. Examining these bottleneck pairs for each k -order distance reveals what are the differences between proximity networks to which d_P^k is most sensitive about. In general, k -order bottleneck pairs tend to be pairs of tuples with high proximity values in their respective networks. The optimal correspondence C^* map tuples with high proximity as closely as possible. Therefore, network distances are typically determined by large proximity values in one of the networks that can't be matched closely to proximity values in the other network.

In the quinquennial coauthorship networks of Figure 5 the bottleneck pair for 0-order distances d_P^0 , is formed by nodes with high zero order proximities and d_P^0 reflects the difference between their zero order proximities. Since the networks are normalized so that the lead nodes have size 1, d_P^0 is determined by their predominant coauthors, i.e., the scholars that collaborated most prolifically with GG or VM during the period of interest. The distances d_P^0 between GG and VM networks are large because these predominant collaborations are different. In GG networks there are usually groups of 3 to 5 predominant collaborators, whereas in MV networks there are usually one or two that concentrate a larger fraction of the total number of publications.

Similarly, high first order proximity distances are likely due to one of the following situations: (i) Large differences between the numbers of papers authored by the predominant collaborators. (ii) Different patterns in the formation of communities – defined here as clusters of pairwise collaboration. In the latter case large distances arise because it is impossible to match the communities in one network to communities in the other. The distances d_P^1 between GG and MV networks are large because the latter contain a smaller number of communities, which are also more strongly connected than the communities in GG networks.

In second order distances the bottleneck pair of triplets may reflect one of the following scenarios: (i) One network has collaboration between four or more authors while the other doesn't. (ii) There exist three authors with a strong collaboration between them in one network whereas in the

other network there does not exist collaboration between three authors or, if such collaboration exists, it is weak. Many papers written by MV are collaborations of three or four scholars and the predominant coauthor in MV networks appears in at least one collaboration of four scholars. For GG, his 2004-2008 network has a few collaborations consisting of four scholars however all such collaborations are weak. His 2009-2013 network has no publications written by four authors.

B. Biennial networks

The networks GG0408 and GG0913 have more nodes than the networks MV0408 and MV0913 prompting the possibility that the differences in distances discussed in Section VI-B are just due to their different number of publications. This is part of the reason, but not all. To see that this is true we consider the biennial GG collaboration networks. Each of these networks contain numbers of papers that are comparable to the number of papers in the quinquennial MV networks.

Two dimensional Euclidean embeddings of the individual k -order distances d_P^k for $k \in \{0, 1, 2\}$ and the aggregate distance $d_{P,1}$ between the 4 quinquennial networks and the 5 biennial networks are shown in Figure 8. An unsupervised classification run across four distances would assign all nine networks correctly (d_P^1, d_P^2) or two of them incorrectly ($d_P^0, d_{P,1}$).

We expect more variation in biennial networks because the time for averaging behavior is reduced. E.g., we may see deviations from usual collaboration patterns due to the presence of exceptional doctoral students. Still, three of the biennial networks, GG0405, GG0607, GG1011, (up triangles) and the two quinquennial networks GG0408, GG0913 (diamonds) are close to each other in every metric used and form a cluster clearly separate from the two five-year networks MV0408 and MV0913 (circles). This is due to the fact that the distinctive features of GG coauthorship are well reflected in GG0405, GG0607, GG1011. These features include: (i) Multiple predominant coauthors, each of whose collaboration with GG does not comprise a dominant portion of GG's scholarship during the period. (ii) Multiple small coauthorship communities in which strong collaborations within each community are rare. (iii) The number of publications with four or more authors is low. These features contrast with the rather opposite properties of the MV networks.

The networks GG0809 and GG1213 (down triangles) do not cluster nicely with the other five GG networks. Depending on

which distance we consider they may be closest to some of the other GG networks or to one of the two MV networks. This is because, likely due to random variation, GG0809 and GG1213 have some features that resemble GG networks and some other features that resemble MV networks. Fundamentally this happens because of the exceptionally prolific collaborations with Ioannis Schizas (IS) in the 2008-2009 period and Gonzalo Mateos (GM) in the 2012-2013 period. In the network GG0809 the IS node commands a significant fraction of GG publications and creates strong links between collaboration clusters that would be otherwise separate. Both of these features are more characteristic of MV networks. In GG1213 network the GM node accounts for half of the publications in which GG is an author. This is, also, a feature more representative of MV networks than of GG networks.

In summary, proximity network distances capture features of scholar collaboration that permit discerning networks of different authors even when we consider networks that have very different numbers of nodes. The zeroth order distance $d_{\mathcal{P}}^0$ responds primarily to the number of predominant coauthors and the proportion of collaboration between predominant coauthors and the central scholar. The first order distance $d_{\mathcal{P}}^1$ is mostly determined by the fraction of collaborations that involve predominant coauthors and the central scholar as well as the level and number of strong collaborations within each community in the group. The second order distance $d_{\mathcal{P}}^2$ is largely given by the existence, level, and number of collaborations between four or more scholars and the appearance of predominant coauthors in a collaboration between four or more scholars.

Remark 3 The proposed metrics successfully identify the distinct collaborative behaviors of Prof. G. B. Giannakis and Prof. M. Vetterli from incomplete subsets of their publication datasets. The distances between Giannakis's networks (either quinquennial or biennial) are smaller than the distances between Giannakis's networks and Vetterli's networks. This proximity can be used in author name disambiguation or related problems, e.g., adjudicate the biennial networks to their rightful author if only the authors of the quinquennial networks are known.

Remark 4 As a comparison, we applied some simple and reasonable methods to compare the corresponding pairwise networks of the coauthorship networks considered in this section. Motifs have been shown effective in distinguishing coauthorship networks from different scientific fields [19]. To compare high order coauthorship networks by motifs, we restrict attention to pairwise relationships. The dissimilarities between coauthorship networks are assigned as the differences between the summations of the weighted motifs in their corresponding pairwise networks. Analysis based on triangle motifs (weighted) results in MV0408, MV0913, GG0408, and GG0809 being closer to each other and GG0913, GG0405, GG0607, GG1011, and GG1213 being more proximate. Tetrahedron motif analysis (weighted) results in MV0408, MV0913, GG0408, GG0405, GG0607, and GG0809 being closer to each other and GG0913, GG1011, and GG1213

being more proximate. Other simple and common methods to compare pairwise networks yield similar results. Methods to compare pairwise networks via features give us similar observations as those based on the metric distances proposed in the paper. Notice that GG0408 and GG0913 are highly similar regarding the proposed network distances however their differences are relatively large in terms of feature comparisons.

VII. CONCLUSION

We have considered high order networks as a generalization of conventional pairwise networks and discussed the definition of valid metrics to enable their comparison. High order networks satisfy the specification of degeneracy relations that relationship function within a tuple of repeating elements is identical to the relationship within its largest subtuple with unique elements. The table in Figure 9 summarizes the results derived in this paper. The fundamental definitions are those of the k -order network differences introduced in Definition 7 and the p -norm difference introduced in Definition 9. Proposition 1 proves that the distances $d_{\mathcal{N}}^k : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ are pseudometrics in the space of networks modulo k -isomorphism. Proposition 2 shows that $d_{\mathcal{N},p} : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ is a pseudometric in the space of networks modulo isomorphism.

We also introduced the space \mathcal{D}^K of dissimilarity networks of order K in Definition 10 and the space \mathcal{P}^K of proximity networks in Definition 13. Dissimilarity networks also satisfy the order increasing property whereby tuples become more dissimilar when members are added to the group. Proximity networks abide to the order decreasing property whereby tuples becomes less similar when adding nodes to the group.

When restricted to the space of dissimilarity networks the distance $d_{\mathcal{D}}^k : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ is termed the k -order dissimilarity network distance [cf. Definition 11] and the distance $d_{\mathcal{D},p} : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ is termed the p -norm dissimilarity network distance [cf. Definition 12]. We proved that the k -order dissimilarity network distance is a metric in the space $\mathcal{D}^K \bmod \cong_k$ of dissimilarity networks modulo k -isomorphism for any integers $k \geq 1$ [cf. Theorem 1] and that the p -norm dissimilarity network distance is a metric in the space $\mathcal{D}^K \bmod \cong$ of dissimilarity networks modulo isomorphism [cf. Theorem 2]. Analogous results hold true for proximity networks as summarized in the last column of the table in Figure 9 and spelled out in Definitions 14 and 15 and Theorems 3 and 4. We have also shown that the p -norm $\|\mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K)\|_p$ of the vector that groups the k -order differences $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ lower bounds the p -norm difference $d_{\mathcal{N},p}(N_X^K, N_Y^K)$ [cf. Proposition 3]. This property is inherited when we restrict attention to proximity and dissimilarity networks as summarized in the bottom row of the table in Figure 9.

Proximity and dissimilarity networks are equivalent constructions as it follows formally from the notion of duality introduced in Definition 16. We have shown that this duality extends to the various distances defined in the sense that proximity distances between two proximity networks is the same as the dissimilarity distances between their corresponding duals [cf. Proposition 4].

Input space	K -order networks, \mathcal{N}^K	K -order dissimilarity networks, \mathcal{D}^K	K -order proximity networks, \mathcal{P}^K
k -order difference	$d_{\mathcal{N}}^k : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ [cf. Definition 7] Pseudometric in $\mathcal{N}^K \bmod \cong_k$	$d_{\mathcal{D}}^k : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ [cf. Definition 11] Metric in $\mathcal{D}^K \bmod \cong_k$, for $k \geq 1$	$d_{\mathcal{P}}^k : \mathcal{P}^K \times \mathcal{P}^K \rightarrow \mathbb{R}_+$ [cf. Definition 14] Metric in $\mathcal{P}^K \bmod \cong_k$, for $k \geq 1$
p -norm difference	$d_{\mathcal{N},p} : \mathcal{N}^K \times \mathcal{N}^K \rightarrow \mathbb{R}_+$ [cf. Definition 9] Pseudometric in $\mathcal{N}^K \bmod \cong$	$d_{\mathcal{D},p} : \mathcal{D}^K \times \mathcal{D}^K \rightarrow \mathbb{R}_+$ [cf. Definition 12] Metric in $\mathcal{D}^K \bmod \cong$	$d_{\mathcal{P},p} : \mathcal{P}^K \times \mathcal{P}^K \rightarrow \mathbb{R}_+$ [cf. Definition 15] Metric in $\mathcal{P}^K \bmod \cong$
Relationships	$d_{\mathcal{N},p}(N_X^K, N_Y^K) \geq \ \mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K)\ _p$	$d_{\mathcal{D},p}(D_X^K, D_Y^K) \geq \ \mathbf{d}_{\mathcal{D}}^K(D_X^K, D_Y^K)\ _p$	$d_{\mathcal{P},p}(P_X^K, P_Y^K) \geq \ \mathbf{d}_{\mathcal{P}}^K(P_X^K, P_Y^K)\ _p$

Fig. 9. Relationships between the spaces of high order networks, dissimilarity networks, and proximity networks. A family of pseudometrics can be defined to measure dissimilarities between a specific order functions between high order networks. Another family of pseudometrics can be defined to quantify distinctions between high order networks across all order functions. These two families of pseudometrics are related and become metrics in the corresponding spaces when we restrict attentions to dissimilarity networks or proximity networks.

We illustrated the value of our definitions by using proximity network distances to successfully identify collaboration patterns of Prof. Georgios B. Giannakis and Prof. Martin Vetterli. With respect to future goals the most important limitation in the current manuscript is that distances are difficult to compute when the number of nodes in the network is large. For networks with large number of nodes it is necessary to develop tools for approximate evaluation of network distances. These tools exist for the comparison of metric spaces and their generalization to networks is part of ongoing research. The idea is to relate high order dissimilarity networks to simplicial complexes and filtrations so that distances between networks can be lower bounded or reasonably approximated by the difference between persistence homologies of the corresponding filtrations [32], [33].

APPENDIX A PROOF OF PROPOSITION 1

To prove that $d_{\mathcal{N}}^k$ for any integer $0 \leq k \leq K$ is a pseudometric in the space of K -order networks modulo k -isomorphism we prove the (i) nonnegativity, (ii) symmetry, (iii') relaxed identity, and (iv) triangle inequality properties in Definition 3.

Proof of nonnegativity property: For any integers $0 \leq k \leq K$, since $|r_X^k(x_{0:k}) - r_Y^k(y_{0:k})|$ is nonnegative $\Gamma_{X,Y}^k(C)$ defined in (8) also is. The network distance must then satisfy $d_{\mathcal{N}}^k(N_X^K, N_Y^K) \geq 0$ because it is a minimum of nonnegative numbers. ■

Proof of symmetry property: A correspondence $C \subseteq X \times Y$ with elements $c_i = (x_i, y_i)$ results in the same associations as the correspondence $\tilde{C} \subseteq Y \times X$ with element $\tilde{c}_i = (y_i, x_i)$. Thus, for any correspondence C and integers $0 \leq k \leq K$, we have a correspondence \tilde{C} such that $\Gamma_{X,Y}^k(C) = \Gamma_{Y,X}^k(\tilde{C})$. It follows that the minima in (9) must coincide from where it follows that $d_{\mathcal{N}}^k(N_X^K, N_Y^K) = d_{\mathcal{N}}^k(N_Y^K, N_X^K)$. ■

Proof of relaxed identity property: We need to show that for any integers $0 \leq k \leq K$ if N_X^K and N_Y^K are k -isomorphic we must have $d_{\mathcal{N}}^k(N_X^K, N_Y^K) = 0$. To see that this is true recall that for k -isomorphic networks there exists a bijection $\phi : X \rightarrow Y$ that preserves distance functions at order k [cf. (7)]. Consider then the particular correspondence $C_\phi = \{(x, \phi(x)), x \in X\}$. For all $x_0 \in X$ there is an element $c = (x_0, y) \in C_\phi$ and for all $y_0 \in Y$ there is an element $c' = (x, y_0) \in C_\phi$ since ϕ is bijective. Thus C_ϕ is a valid

correspondence between X and Y for which (7) indicates that it must be

$$r_Y^k(y_{0:k}) = r_Y^k(\phi(x_{0:k})) = r_X^k(x_{0:k}), \quad (26)$$

for any $(x_{0:k}, y_{0:k}) \in C_\phi$. This implies $\Gamma_{X,Y}^k(C) = |r_X^k(x_{0:k}) - r_Y^k(y_{0:k})| = 0$ for any $(x_{0:k}, y_{0:k}) \in C_\phi$. Since C_ϕ is a particular correspondence, taking a minimum over all correspondences as in (9) yields

$$d_{\mathcal{N}}^k(N_X^K, N_Y^K) \leq \Gamma_{X,Y}^k(C) = 0. \quad (27)$$

Since $d_{\mathcal{N}}^k(N_X^K, N_Y^K) \geq 0$, as already shown, it must be that $d_{\mathcal{N}}^k(N_X^K, N_Y^K) = 0$ when N_X^K and N_Y^K are k -isomorphic. ■

Proof of triangle inequality: To show that the triangle inequality holds, let the correspondence C_1 between X and Z and the correspondence C_2 between Z and Y be the minimizing correspondences in (9). We can then write

$$d_{\mathcal{N}}^k(N_X^K, N_Z^K) = \Gamma_{X,Z}^k(C_1), \quad d_{\mathcal{N}}^k(N_Z^K, N_Y^K) = \Gamma_{Z,Y}^k(C_2). \quad (28)$$

Define a correspondence C between X and Y as the one induced by pairs (x, z) and (z, y) sharing a common node $z \in Z$,

$$C := \{(x, y) \mid \exists z \in Z \text{ with } (x, z) \in C_1, (z, y) \in C_2\}. \quad (29)$$

To show that C is a well defined correspondence we need to show that for every $x \in X$ there exists $y_0 \in Y$ such that $(x, y_0) \in C$ and by symmetry for every $y \in Y$ there exists $x_0 \in X$ such that $(x_0, y) \in C$. To see this, first pick an arbitrary $x \in X$. Because C_1 is a correspondence between X and Z there must exist $z_0 \in Z$ such that $(x, z_0) \in C_1$. There must exist $y_0 \in Y$ such that $(z_0, y_0) \in C_2$ since C_2 is also a correspondence between Y and Z . Therefore, there exists a pair $(x, y_0) \in T$ with $y_0 \in Y$ for any $x \in X$. The second part follows by symmetry and C is a well defined correspondence. The correspondence C may not be the minimizing correspondence for the distance $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$. However since it is a valid correspondence with the definition in (9) we can write

$$d_{\mathcal{N}}^k(N_X^K, N_Y^K) \leq \Gamma_{X,Y}^k(C). \quad (30)$$

By the definition of C in (29), the requirement $(x_{0:k}, y_{0:k}) \in C$ is equivalent as $(x_{0:k}, z_{0:k}) \in C_1$ and $(z_{0:k}, y_{0:k}) \in C_2$ for any $0 \leq k \leq K$. Further adding and subtracting $r_Z^k(z_{0:k})$ in the

absolute value of $\Gamma_{X,Y}^k(C) = |r_X^k(x_{0:k}) - r_Y^k(y_{0:k})|$ and using the triangle inequality of the absolute value yields

$$\Gamma_{X,Y}^k(C) \leq \max_{\substack{(x_{0:k}, z_{0:k}) \in C_1 \\ (z_{0:k}, y_{0:k}) \in C_2}} \left\{ |r_X^k(x_{0:k}) - r_Z^k(z_{0:k})| + |r_Z^k(z_{0:k}) - r_Y^k(y_{0:k})| \right\}. \quad (31)$$

We can further bound (31) by taking maximum over each summand,

$$\Gamma_{X,Y}^k(C) \leq \max_{(x_{0:k}, z_{0:k}) \in C_1} |r_X^k(x_{0:k}) - r_Z^k(z_{0:k})| + \max_{(z_{0:k}, y_{0:k}) \in C_2} |r_Z^k(z_{0:k}) - r_Y^k(y_{0:k})| = \Gamma_{X,Z}^k(C_1) + \Gamma_{Z,Y}^k(C_2). \quad (32)$$

Substituting (30) and (28) into (32) yields triangle inequality. ■

APPENDIX B PROOF OF PROPOSITION 2

To prove that $d_{\mathcal{N},p}$ is a distance in the space of K -order networks modulo isomorphism we prove the (i) nonnegativity, (ii) symmetry, (iii') relaxed identity, and (iv) triangle inequality properties in Definition 3.

Proof of nonnegativity property: Since $\|\Gamma_{X,Y}^k(C)\|_p \geq 0$, the network distance must then satisfy $d_{\mathcal{N},p}(N_X^K, N_Y^K) \geq 0$ as it is a minimum of nonnegative numbers. ■

Proof of symmetry property: A correspondence $C \subseteq X \times Y$ with elements $c_i = (x_i, y_i)$ results in the same associations as the correspondence $\tilde{C} \subseteq Y \times X$ with element $\tilde{c}_i = (y_i, x_i)$. Thus, for any correspondence C we have a correspondence \tilde{C} such that $\Gamma_{X,Y}^k(C) = \Gamma_{Y,X}^k(\tilde{C})$. This implies $\|\Gamma_{X,Y}^k(C)\|_p = \|\Gamma_{Y,X}^k(\tilde{C})\|_p$. It follows that the minima in (11) must coincide and therefore $d_{\mathcal{N},p}(N_X^K, N_Y^K) = d_{\mathcal{N},p}(N_Y^K, N_X^K)$. ■

Proof of relaxed identity property: We need to show that if N_X^K and N_Y^K are isomorphic we must have $d_{\mathcal{N},p}(N_X^K, N_Y^K) = 0$. To see that this is true recall that for isomorphic networks there exists a bijection $\phi : X \rightarrow Y$ that preserves distance functions at every order [cf. (7)]. Consider then the particular correspondence $C_\phi = \{(x, \phi(x)), x \in X\}$. We have demonstrated in Appendix A that C_ϕ is a valid correspondence between X and Y . The definition of isomorphism indicates that it must be (26) holds true for all $0 \leq k \leq K$ and $(x_{0:k}, y_{0:k}) \in C_\phi$. Since C_ϕ is a particular correspondence, from (11) it follows that

$$d_{\mathcal{N},p}(N_X^K, N_Y^K) \leq \|\Gamma_{X,Y}^k(C_\phi)\|_p. \quad (33)$$

Because $r_X^k(x_{0:k}) - r_Y^k(y_{0:k}) = 0$ for any $0 \leq k \leq K$ and any $(x_{0:k}, y_{0:k}) \in C_\phi$ by (26), we have $\Gamma_{X,Y}^k(C_\phi) = 0$. $\|\cdot\|_p$ being a proper norm implies $\|\Gamma_{X,Y}^k(C_\phi)\|_p = 0$. Substituting this back into (33) shows $d_{\mathcal{N},p}(N_X^K, N_Y^K) \leq 0$. Since $d_{\mathcal{N},p}(N_X^K, N_Y^K) \geq 0$, as already shown, it must be that $d_{\mathcal{N},p}(N_X^K, N_Y^K) = 0$ when N_X^K and N_Y^K are isomorphic. ■

Proof of triangle inequality: To show that the triangle inequality holds, let the correspondence C_1 between X and

Z and the correspondence C_2 between Z and Y be the minimizing correspondences in (11). We can then write

$$\begin{aligned} d_{\mathcal{N},p}(N_X^K, N_Z^K) &= \|\Gamma_{X,Z}^k(C_1)\|_p, \\ d_{\mathcal{N},p}(N_Z^K, N_Y^K) &= \|\Gamma_{Z,Y}^k(C_2)\|_p. \end{aligned} \quad (34)$$

Define a correspondence C between X and Y in the same way as (29). We have demonstrated in Appendix A that C is a well defined correspondence. Therefore with the definition in (11) we can write

$$d_{\mathcal{N},p}(N_X^K, N_Y^K) \leq \|\Gamma_{X,Y}^k(C)\|_p. \quad (35)$$

Moreover, in Appendix A we also showed for any $0 \leq k \leq K$,

$$\Gamma_{X,Y}^k(C) \leq \Gamma_{X,Z}^k(C_1) + \Gamma_{Z,Y}^k(C_2). \quad (36)$$

This implies the vector $\Gamma_{X,Z}^k(C_1) + \Gamma_{Z,Y}^k(C_2)$ is elementwise no smaller than the vector $\Gamma_{X,Y}^k(C)$. The definition of p -norm $\|\mathbf{x}\|_p = (\sum_{k=0}^K |x_i|^p)^{1/p}$ guarantees that the value of $\|\mathbf{x}\|_p$ is monotonically nondecreasing on each element x_i in $\mathbf{x} = (x_0, x_1, \dots, x_n)^T$. Therefore,

$$\|\Gamma_{X,Y}^k(C)\|_p \leq \|\Gamma_{X,Z}^k(C_1) + \Gamma_{Z,Y}^k(C_2)\|_p. \quad (37)$$

We can further bound (37) by using the triangle inequality of the p -norm,

$$\|\Gamma_{X,Y}^k(C)\|_p \leq \|\Gamma_{X,Z}^k(C_1)\|_p + \|\Gamma_{Z,Y}^k(C_2)\|_p. \quad (38)$$

Substituting (35) and (34) back into (38) yields the triangle inequality. ■

APPENDIX C PROOFS IN SECTION IV-A

Proof of Theorem 1: The proof in Appendix A has demonstrated $d_{\mathcal{D}}^k$ is a pseudometric in the space $\mathcal{D}^K \text{ mod } \cong_k$. To prove that $d_{\mathcal{D}}^k$ is a metric in the same space we need to show the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: We want to prove $d_{\mathcal{D}}^k(D_X^K, D_Y^K) = 0$ must imply that D_X^K and D_Y^K are k -isomorphic. If $d_{\mathcal{D}}^k(D_X^K, D_Y^K) = 0$, there exists a correspondence C such that $r_X^k(x_{0:k}) = r_Y^k(y_{0:k})$ for any $(x_{0:k}, y_{0:k}) \in C$. Define a function $\phi : X \rightarrow Y$ that associates x with an arbitrary y chosen from the set that form a pair with x in C ,

$$\phi : x \mapsto y_0 \in \{y \mid (x, y) \in C\}. \quad (39)$$

Since C is a correspondence the set $\{y \mid (x, y) \in C\}$ is nonempty for any x implying that ϕ is well-defined for any $x \in X$. Therefore $r_X^k(x_{0:k}) = r_Y^k(\phi(x_{0:k}))$ for any $x_{0:k}$. This implies the function ϕ must be injective. If it were not, there would be a pair of nodes $x \neq x'$ with $\phi(x) = \phi(x') = y$ for some $y \in Y$. Hence the k -order relationship function between (x, \dots, x, x') where the first $k-1$ nodes in the tuple are x and the last node is x' would satisfy

$$r_X^k(x, \dots, x, x') = r_Y^k(\phi(x, \dots, x, x')) = r_Y^k(y, \dots, y), \quad (40)$$

follows from the definition of ϕ . The k -order relationship between the tuple (x, \dots, x) where all the k nodes are identical would also satisfy

$$r_X^k(x, \dots, x) = r_Y^k(\phi(x, \dots, x)) = r_Y^k(y, \dots, y). \quad (41)$$

Combining (40) and (41) yields

$$r_X^k(x, \dots, x, x') = r_X^k(x, \dots, x). \quad (42)$$

Meanwhile, the identity property for high order networks [cf. Definition 5] implies

$$r_X^k(x, \dots, x, x') = r_X^2(x, x'), \quad r_X^k(x, \dots, x) = r_X^1(x). \quad (43)$$

Using the fact that for dissimilarity networks, relationship functions are the summations of dissimilarity functions and the multiplication of ϵ and ranks, we have that

$$r_X^2(x, x') = d_X^2(x, x') + 2\epsilon, \quad r_X^1(x) = d_X^1(x) + \epsilon. \quad (44)$$

Moreover, the order increasing property for dissimilarity functions imply

$$d_X^2(x, x') \geq d_X^1(x). \quad (45)$$

Substituting the decompositions (44) and (45) into (43) yields

$$r_X^k(x, \dots, x, x') > r_X^k(x, \dots, x). \quad (46)$$

which contradicts with (42) and shows that ϕ must be injective.

Likewise, define the function $\psi : Y \rightarrow X$ that associates y with an arbitrary x chosen from the set that form a pair with y in C ,

$$\psi : y \mapsto x_0 \in \{x | (x, y) \in C\}. \quad (47)$$

It follows by similar arguments that ψ must be injective. By applying the Cantor-Bernstein-Schroeder theorem [34, Section 2.6] to the reciprocal injections $\phi : X \rightarrow Y$ and $\psi : Y \rightarrow X$, the existence of a bijection between X and Y is guaranteed. This forces X and Y to have same cardinality and ϕ and ψ being bijections. Pick the bijection ϕ and it follows $r_X^k(x_{0:k}) = r_Y^k(\phi(x_{0:k}))$ for all nodes $(k+1)$ -tuples $x_{0:k} \in X^{k+1}$. This shows that $D_X^K \cong_k D_Y^K$ and completes the proof of the identity statement. ■

Having demonstrated all four properties in Theorem 1, the global proof completes. ■

Proof of Theorem 2: The proof in Appendix B has demonstrated that $d_{\mathcal{D},p}$ is a pseudometric in the space $\mathcal{D}^K \bmod \cong$. To prove that $d_{\mathcal{D},p}$ is a metric in the same space we further demonstrate the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: We want to show $d_{\mathcal{D},p}(D_X^K, D_Y^K) = 0$ implying D_X^K and D_Y^K being isomorphic. If $d_{\mathcal{D},p}(D_X^K, D_Y^K) = \min_{C \in \mathcal{C}(X,Y)} \|\Gamma_{X,Y}^K(C)\|_p = 0$, there exists a correspondence C such that

$$\|\Gamma_{X,Y}^K(C)\|_p = 0. \quad (48)$$

The property of p -norm implies that this correspondence C satisfies $\Gamma_{X,Y}^k(C) = 0$ for $0 \leq k \leq K$, i.e. $r_X^k(x_{0:k}) = r_Y^k(y_{0:k})$ for any $0 \leq k \leq K$ and $(x_{0:k}, y_{0:k}) \in C$. Define

functions $\phi : X \rightarrow Y$ as in (39) and $\psi : Y \rightarrow X$ as in (47). The analysis in Proof of Theorem 1 has demonstrated that ϕ and ψ are bijections and that X and Y have same cardinality. Pick the bijection ϕ and it follows $r_X^k(x_{0:k}) = r_Y^k(\phi(x_{0:k}))$ for any $0 \leq k \leq K$ and all $(k+1)$ -tuples $x_{0:k} \in X$. This shows that $D_X^K \cong D_Y^K$ and completes the proof of the identity statement. ■

APPENDIX D PROOFS IN SECTION V-A

Proof of Theorem 3 : The proof in Appendix A has demonstrated that $d_{\mathcal{P}}^k$ is a pseudometric in the space $\mathcal{P}^K \bmod \cong_k$. To prove that $d_{\mathcal{P}}^k$ is a metric in the same space we need to show the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: Most parts of the proof follow from the proof of the second part of the identity property for Theorem 1 in Appendix C. The only difference is in demonstrating the function ϕ constructed in (39) is injective. Under the same setup where there exist a pair of nodes $x \neq x'$ such that $\phi(x) = \phi(x') = y$ for some $y \in Y$, the k -order relationship between (x, \dots, x, x') would satisfy

$$r_X^k(x, \dots, x, x') = r_Y^k(y, \dots, y) = r_X^k(x, \dots, x). \quad (49)$$

Meanwhile, the facts of proximities in proximity networks follow order decreasing property $p_X^2(x, x') \leq p_X^1(x)$ and $r_X^2(x, x') = p_X^2(x, x') - 2\epsilon, r_X^1(x) = p_X^1(x) - \epsilon$ from (19) implies

$$r_X^2(x, x') < r_X^1(x). \quad (50)$$

Combining (50) with the identity property inherited from high order networks [cf. Definition 5] $r_X^k(x, \dots, x, x') = r_X^2(x, x'), r_X^k(x, \dots, x) = r_X^1(x)$ gives us

$$r_X^k(x, \dots, x, x') < r_X^k(x, \dots, x), \quad (51)$$

which contradicts with (49) and shows that ϕ must be injective. The rest of the proof follows. ■

Proof of Theorem 4: The proof in Appendix B has demonstrated that $d_{\mathcal{P},p}$ is a pseudometric in the space $\mathcal{P}^K \bmod \cong$. To prove that $d_{\mathcal{P},p}$ is a metric in the same space we further demonstrate the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: We want to show that having $d_{\mathcal{P},p}(P_X^K, P_Y^K) = 0$ must imply that P_X^K being isomorphic to P_Y^K . If $d_{\mathcal{P},p}(P_X^K, P_Y^K) = 0$, there exists a correspondence C such that $\|\Gamma_{X,Y}^K(C)\|_p = 0$. The property of p -norm implies that this correspondence C satisfies $r_X^k(x_{0:k}) = r_Y^k(y_{0:k})$ for any $0 \leq k \leq K$ and any $(x_{0:k}, y_{0:k}) \in C$. Define functions $\phi : X \rightarrow Y$ as in (39) and $\psi : Y \rightarrow X$ as in (47), the analysis in Appendix D Proof of Theorem 3 has demonstrated that ϕ and ψ are bijections and that X and Y have same cardinality. Pick the bijection ϕ and it follows $r_X^k(x_{0:k}) = r_Y^k(\phi(x_{0:k}))$ for any $0 \leq k \leq K$ and $x_{0:k} \in X$. This shows that $P_X^K \cong P_Y^K$ and completes the proof of the identity statement. ■

APPENDIX E
PROOFS IN SECTION V-B

Proof of Proposition 4 : We first prove (23) by considering proximity networks P_X^K and P_Y^K and their corresponding dual dissimilarity networks D_X^K and D_Y^K . Let the correspondence C between X and Y be the minimizing correspondence in $d_P^k(P_X^K, P_Y^K)$ [cf. Definition 14] so that we can write

$$d_P^k(P_X^K, P_Y^K) = \Gamma_{P_X, P_Y}^k(C). \quad (52)$$

C may not be the minimizing correspondence for the distance $d_D^k(D_X^K, D_Y^K)$ [cf. Definition 11], but since it is a valid correspondence, it holds true that

$$d_D^k(D_X^K, D_Y^K) \leq \Gamma_{D_X, D_Y}^k(C). \quad (53)$$

From the definition of duality [cf. (22)], we may write

$$\Gamma_{D_X, D_Y}^k(C) = \max_{(x_{0:k}, y_{0:k}) \in C} \left| (1 - \hat{d}_X^k(x_{0:k})) - (1 - \hat{d}_Y^k(y_{0:k})) \right|. \quad (54)$$

The ones in (54) cancel out and therefore,

$$\Gamma_{D_X, D_Y}^k(C) = \Gamma_{P_X, P_Y}^k(C). \quad (55)$$

Substituting (52) and (53) back to (55) implies

$$d_P^k(P_X^K, P_Y^K) \geq d_D^k(D_X^K, D_Y^K). \quad (56)$$

Let the correspondence C' between X and Y be the minimizing correspondence in $d_D^k(D_X^K, D_Y^K)$. Then C' is also a valid correspondence for the distance $d_P^k(P_X^K, P_Y^K)$. By symmetry, we have

$$d_D^k(D_X^K, D_Y^K) \geq d_P^k(P_X^K, P_Y^K). \quad (57)$$

Combining (56) and (57) yields the desired result in (23).

Next we prove (24) by considering P_X^K and P_Y^K and their corresponding duals D_X^K and D_Y^K . Let the correspondence C between X and Y be the minimizing correspondence in $d_{P,p}(P_X^K, P_Y^K)$ [cf. Definition 12] so that we can write

$$d_{P,p}(P_X^K, P_Y^K) = \|\Gamma_{P_X, P_Y}^k(C)\|_p. \quad (58)$$

C may not be the minimizing correspondence for the distance $d_{D,p}(D_X^K, D_Y^K)$ [cf. Definition 12], but again since it is a valid correspondence, we may write

$$d_{D,p}(D_X^K, D_Y^K) \leq \|\Gamma_{D_X, D_Y}^k(C)\|_p. \quad (59)$$

We have demonstrated in proving (23) that for any integers $0 \leq k \leq K$, $\Gamma_{D_X, D_Y}^k(C) = \Gamma_{P_X, P_Y}^k(C)$. In vector form, this is $\Gamma_{D_X, D_Y}^K(C) = \Gamma_{P_X, P_Y}^K(C)$. Therefore, the property of p -norm implies that

$$\|\Gamma_{D_X, D_Y}^K(C)\|_p = \|\Gamma_{P_X, P_Y}^K(C)\|_p. \quad (60)$$

Substituting (58) and (59) back to (60) yields

$$d_{P,p}(P_X^K, P_Y^K) \geq d_{D,p}(D_X^K, D_Y^K). \quad (61)$$

Let the correspondence C' between X and Y be the minimizing correspondence in $d_{D,p}(D_X^K, D_Y^K)$. Then C' is also a valid correspondence for $d_{P,p}(P_X^K, P_Y^K)$. By symmetry, we have

$$d_{D,p}(D_X^K, D_Y^K) \geq d_{P,p}(P_X^K, P_Y^K). \quad (62)$$

Combining (61) and (62) yields the desired result in (24). ■

REFERENCES

- [1] W. Huang and A. Ribeiro, "Metrics in the Space of High Order Proximity Networks," in *Proc. Int. Conf. Acoustics Speech Signal Process*, Brisbane, Australia, April 19-24 2015, pp. 4135-4139.
- [2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., 1993.
- [3] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, ser. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.
- [4] R. Ghrist and A. Muhammad, "Coverage and hole-detection in sensor networks via homology," in *International Symposium on Information Processing in Sensor Networks*, vol. 00, no. 1, 2005, pp. 254-260.
- [5] V. de Silva and R. Ghrist, "Coordinate-free Coverage in Sensor Networks with Controlled Boundaries via Homology," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1205-1222, Dec. 2006.
- [6] H. Chintakunta and H. Krim, "Divide and Conquer: Localizing Coverage Holes in Sensor Networks," in *2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, Jun. 2010, pp. 1-8.
- [7] B. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory," *Computational Intelligence Magazine, IEEE*, vol. 3, no. 3, pp. 49-63, 2008.
- [8] W. Ren, Q. Zhao, R. Ramanathan, J. Gao, A. Swami, A. Bar-Noy, M. P. Johnson, and P. Basu, "Broadcasting in multi-radio multi-channel wireless networks using simplicial complexes," in *Wireless Networks*, vol. 19, no. 6, Nov. 2012, pp. 1121-1133.
- [9] J. Xu and V. Singh, "Unified Hypergraph for Image Ranking in a Multimodal Context," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 2333-2336.
- [10] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4290-303, Sep. 2012.
- [11] A. Wilkerson, T. Moore, A. Swami, and H. Krim, "Simplifying the Homology Of Networks via Strong Collapses," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 5258-5262.
- [12] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 18, pp. 7332-6, May 2007.
- [13] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *Science (New York, N.Y.)*, vol. 311, no. 5757, pp. 88-90, Jan. 2006.
- [14] S. Segarra, M. Eisen, and A. Ribeiro, "Authorship attribution using function words adjacency networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, no. 2, 2013, pp. 5563-5567.
- [15] D. Khmelev and F. Tweedie, "Using Markov Chains for Identification of Writer," *Literary and linguistic computing*, vol. 16, no. 3, 2001.
- [16] T. Wang and H. Krim, "Statistical classification of social networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 3977-3980.
- [17] R. Singh, J. Xu, and B. Berger, "Global alignment of multiple protein interaction networks with application to functional orthology detection," *Proceedings of the National Academy of Sciences*, vol. 105, no. 35, pp. 12763-12768, 2008.
- [18] L. Peng, L. Liu, S. Chen, and Q. Sheng, "A network comparison algorithm for predicting the conservative interaction regions in protein-protein interaction network," in *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, Sep. 2010, pp. 34-39.
- [19] S. Choobdar, P. Ribeiro, S. Bugla, and F. Silva, "Comparison of Co-authorship Networks across Scientific Fields Using Motifs," *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 147-152, Aug. 2012.
- [20] L. Yong, Z. Yan, and C. Lei, "Protein-protein interaction network comparison based on wavelet and principal component analysis," in *2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, 2010, pp. 430-437.
- [21] N. Pržulj, "Biological network comparison using graphlet degree distribution," *Bioinformatics*, vol. 23, no. 2, pp. e177-183, Jan. 2007.
- [22] T. Milenković and N. Pržulj, "Uncovering biological network function via graphlet degree signatures," *Cancer informatics*, vol. 6, p. 257, Jan. 2008.

- [23] N. Shervashidze, S. Vishwanathan, T. H. Petri, K. Mehlhorn, and K. M. Borgwardt, "Efficient graphlet kernels for large graph comparison," in *International Conference on Artificial Intelligence and Statistics*, vol. 5, 2009, pp. 488–495.
- [24] G. Carlsson, F. Memoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," 2014. [Online]. Available: <http://arxiv.org/abs/1301.7724>
- [25] S. Fortin, "The graph isomorphism problem," Ph.D. dissertation, The University of Alberta, Edmonton, Alberta, Canada, 1996.
- [26] M. Gromov, *Metric structures for Riemannian and non- Riemannian spaces*. Birkha user Boston Inc., Boston, MA., 2007.
- [27] F. Memoli, "Gromov-Hausdorff distances in Euclidean spaces," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2008, pp. 1–8.
- [28] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*. American Mathematical Soc., 2001, vol. 33.
- [29] C. Berge, *Graphs and Hypergraphs*. North-Holland Publishing Company, 1976.
- [30] A. Bretto, *Hypergraph Theory: an Introduction*. Springer, 2013.
- [31] "Engineering Village: the place to find answers to engineering questions." [Online]. Available: <http://www.engineeringvillage.com/search/quick.url>
- [32] W. Huang and A. Ribeiro, "Persistent Homology Approximations of Network Distances," in *Proc. Global Conf. Signal Info. Process.*, vol. (to appear), Orlando FA, Dec 14-16 2015.
- [33] W. Huang and A. Ribeiro, "Persistent homology lower bounds on high order network distances," *Signal Processing, IEEE Transactions on*, vol. (submitted), 2015. [Online]. Available: <http://arxiv.org/abs/1507.03044>
- [34] A. N. Kolmogorov, S. V. Fomine, and R. A. Silverman, *Introductory Real Analysis*. Dover, 1975.